

FESTSCHRIFT SYMPOSIUM FOR

WALTER C. LABYS

**Agricultural and Resource Economics
West Virginia University**

May 7, 2007

Dynamic Quadratic Programming in Process Control

Bruce A Bancroft
1/8/07

CONSOL Energy Inc.
CNX Coal Operations
1800 Washington Road
Pittsburgh, PA 15241-1421

This is a preliminary draft and should not be quoted or reproduced

Dynamic Quadratic Programming in Process Control

In the coal industry real-time blending is in many cases required to satisfy contractual requirements on coal quality, e.g., ash. Continuous nuclear analyzers are used to obtain real time feedback of the quality of the blended product and produce a quality generally at a rate of one per minute. The product typically consists of two sources of coal one which has a quality higher than the target quality and one which is lower than the target quality. The object is to control the proportions of the two sources to achieve the target quality while maximizing the usage of the higher quality which is typically a lower cost source. This presents some problems because the two sources are each autocorrelated in time, the proportions of the sources are measured with error, and the blended product quality measured by the analyzer although measured without bias is measured with significant error. In this paper a methodology developed and implemented to blend coal to meet a barge specification is described.

In process control the work horse methodology for feedback control is the Proportional-Integral-Derivative(PID) controller. Mathematically it is given by:

$$Output(t) = K_p e(t) + K_i \int_0^t e(t) dt + K_d \frac{de}{dx} \quad (1)$$

where,

$$e(t) = SP - PV(t) \quad (2)$$

and SP is the setpoint(desired quality) and PV(t) is the process variable, i.e, the value measured by the continuous analyzer.

In practice generally a P only or a PI controller is implemented, i.e., just the first or the first two terms in (1). The first term is proportional in that the control action is proportional to the deviation between the SP and the PV. Sometimes this is good enough even though it can be shown that it will result in a sustained bias, ie, it never quite gets there much like if you move from one side of a room to the other moving in increments of half the distance remaining. To correct for this bias frequently the integral term is implemented resulting in a control action proportional to the sum of the errors thus eliminating steady state errors. The bad news is it also guarantees overshoot since the integral value will continue to be added to the output value. Derivative control, can be useful in many instances, however in the presence of noise, one can end up chasing this noise and thus inducing unwanted oscillations in the process variable. Detailed descriptions of PID controllers can be found in most control engineering textbooks[1,2].

Implementation of a P controller was attempted in this application with mixed results. Its' performance is satisfactory if the control period is long and what is desired is that the average quality over the time period hits the target. If the period is short the performance is not acceptable.

An alternative approach was investigated since sufficient information exists to optimally forecast the impact a control action will have on the process variable in real time. This approach, in the control literature, is called model predictive control and has become very popular over the last twenty years especially in the chemical processing industry [3,4].

The first method attempted was to implement ordinary least squares in order to disaggregate the product ash into the two sources. The estimates could then be used, given the quality of coal already loaded on the barge, to set blend proportions for the coal to be added so that the final barge quality would meet specifications, while at the same time maximizing the amount of raw coal loaded.

Each signal from the analyzer can be written as:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \quad (3)$$

where,

y_t = the analyzer signal in tons of ash at time t

β_1 = clean ash fraction at time t

x_{1t} = tons of clean coal at time t

β_2 = raw ash fraction at time t

x_{2t} = tons of raw coal at time t

ε_t = analytical error from analyzer at time t

A set of T signals can be represented by the following system of equations:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

where,

\mathbf{y} is a Tx1 vector of the observed ash signals in tons of ash

\mathbf{x} is a Tx2 matrix of source tons

$\boldsymbol{\beta}$ is a 2x1 vector of the clean and raw coal ashes, and

$\boldsymbol{\varepsilon}$ is a Tx1 vector of errors.

The sum of squared errors, viewed as a function of the regression coefficients, is given by:

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) \quad (5)$$

$$= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{x}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{x}'\mathbf{x}\boldsymbol{\beta} \quad (6)$$

$$= \sum_{t=1}^T [y_t - (\beta_1 x_{1t} + \beta_2 x_{2t})]^2$$

The estimates of β will be obtained by finding $\beta \in D \subset R^2$ such that β minimizes $\mathcal{E}'\mathcal{E}$. This is just a standard OLS regression formulation of the problem and the objective function to be minimized is (5) above, which is a polynomial of order 2 in the coefficients of β . Since $\mathbf{y}'\mathbf{y}$ is a constant then minimizing

$$k = -2\mathbf{y}'\mathbf{x}\beta + \beta'\mathbf{x}'\mathbf{x}\beta \quad (7)$$

is equivalent to minimizing (5) above.

The OLS formulation has both pitfalls and shortcomings. A major pitfall is that for many sets of T signals the independent variables will not change significantly and when they do change total tonnage is kept constant. Thus the two independent variables will be almost linearly related (whether in tons or fraction of total) producing a large instability in the estimated regression coefficients. This instability is seen in the Figure 1 where OLS regression was performed on product ash for 16 hours in order to resolve the ash into its clean and raw component qualities. The regressions were of size $N = 25$ and updated every minute with the most recent 25 observations and what is displayed is the condition number for each regression. The condition number is the ratio of the largest eigenvalue to the smallest calculated on $X'X$.

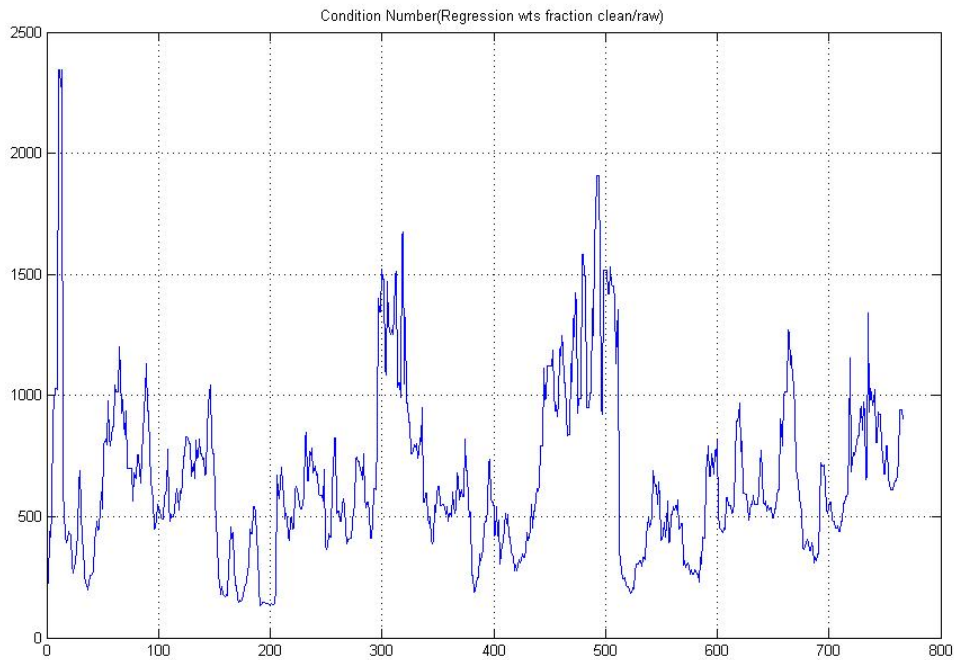


Figure 1 Condition Number (Regression Size = 25) for 16 hrs

Certainly multicollinearity in the regressors is a problem and estimates achieved from OLS correspondingly were highly variable and in many cases not sensible. Another shortfall of OLS is that it does not incorporate all the information that is available for forecasting. Firstly it is known that raw ash is always greater than clean ash and that the clean ash can be bounded. Secondly coal quality is spatially autocorrelated in situ. This

spatial autocorrelation in the ground manifests itself in a temporal autocorrelation in both raw plant feed and clean plant product and thus in the raw/clean blend. It is known that the temporal correlation is non-stationary, this non-stationarity resulting from what mining faces are running at any point in time, where they are mining (and they move), and due to varying mining conditions on any one mining face[5].

A way to incorporate a priori knowledge about the raw and clean ashes is via the constraints methods of quadratic programming where the objective function to minimize is the error sum of squares. The error sum of squares is a quadratic function of the regression coefficients and the optimization can be performed subject to linear inequality constraints.

A way to explicitly accommodate autocorrelation is via generalized (weighted) least squares where the weighting matrix contains the autocorrelation coefficients[6]. If the observations are weighted, then the objective function is the weighted sum of squared errors:

$$(\mathbf{W}\boldsymbol{\varepsilon})'(\mathbf{W}\boldsymbol{\varepsilon}) = [\mathbf{W}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})]'[\mathbf{W}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})] \quad (8)$$

Where, \mathbf{W} is a $T \times T$ diagonal matrix with the ones on the diagonal and the off diagonals containing the autocorrelation coefficients:

$$\mathbf{W} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{pmatrix} \quad (9)$$

In this case the covariance matrix is $\sigma^2\mathbf{W}$. Unfortunately, the temporal correlation is non-stationary. This non-stationarity arises from non-stationarity in in-situ ash, variations in the percent contribution of the different mining sections to mine product, movement of mining sections over time, and due to varying mining conditions on any one mining face. To dynamically model and estimate the autocorrelational structure would add a layer of complexity to the real time algorithm that it would be better to avoid.

If all we care about is a forecast and not standard errors on the regression estimates or prediction intervals but want to somehow accommodate autocorrelation then a simple weighting matrix can be used for this purpose. Furthermore, while there is considerable variation in raw ash at a short scale about a mean level which may be drifting, there is a clear need for protection from gross changes in the raw ash level which does occur and can be traced to changes in how the coal is being mined and/or loaded. If data is autocorrelated and protection is desired from abrupt shifts in raw ash level then more

recent information is more important for estimating purposes, i.e., should carry more weight than older data. If the weighting matrix takes the form of:

$$\mathbf{W} = \begin{pmatrix} w_1^{1/2} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_T^{1/2} \end{pmatrix} \quad (10)$$

then each error in the objective function is being weighted by w_t as shown below.

Expanding (8) the weighted objective function can be written as

$$\boldsymbol{\varepsilon}' \mathbf{W}^2 \boldsymbol{\varepsilon} = \mathbf{y}' \mathbf{W}^2 \mathbf{y} - 2\mathbf{y}' \mathbf{W}^2 \mathbf{x} \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{x}' \mathbf{W}^2 \mathbf{x} \boldsymbol{\beta} \quad (11)$$

$$= \sum_{t=1}^T w_t \left[y_t - (\beta_1 x_{1t} + \beta_2 x_{2t}) \right]^2 \quad (12)$$

The weights are generated by the following function where T is the regression size and the weight assigned to the oldest error is given by w_1 and the weight assigned to the most recent error w_T :

$$w_t = \lambda(1 - \lambda)^{T-t} / \sum_1^T \lambda(1 - \lambda)^{T-t} \quad (13)$$

for $t = 1, \dots, T$ and $0 \leq \lambda \leq 1$. Generally, by experience, optimal lambdas are in the interval [.1,.4]. The denominator in (13) normalizes the weights to sum to one. Thus for a regression of size 25 the unnormalized weight for the most recent observation is λ , since $T-t = 0$, and the unnormalized weight assigned to the oldest observation is $\lambda(1 - \lambda)^{25}$. Note that these are just the weights used to generate an exponentially weighted moving average.

Since $\mathbf{y}' \mathbf{W}^2 \mathbf{y}$ is a constant the QP problem is formulated as Minimize k, where:

$$k = -2\mathbf{y}' \mathbf{W}^2 \mathbf{x} \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{x}' \mathbf{W}^2 \mathbf{x} \boldsymbol{\beta} \quad (14)$$

Subject to:

$$A_C < A_C(\max) \quad (15)$$

$$A_C > A_C(\min) \quad (16)$$

$$A_R > A_C \quad (17)$$

and, A_C is clean ash; A_R is raw ash.

An example illustrating the advantage of weighting the observations is shown via simulation in Figure 2. In this case a shift in raw ash from a mean level of 22% to an average of 30% was simulated. The top plot shows the estimated raw ash against the simulated actual using a QP where the observations are not weighted. The second plot shows the impact of weighting the observations and as shown recovered from the shift much more rapidly.

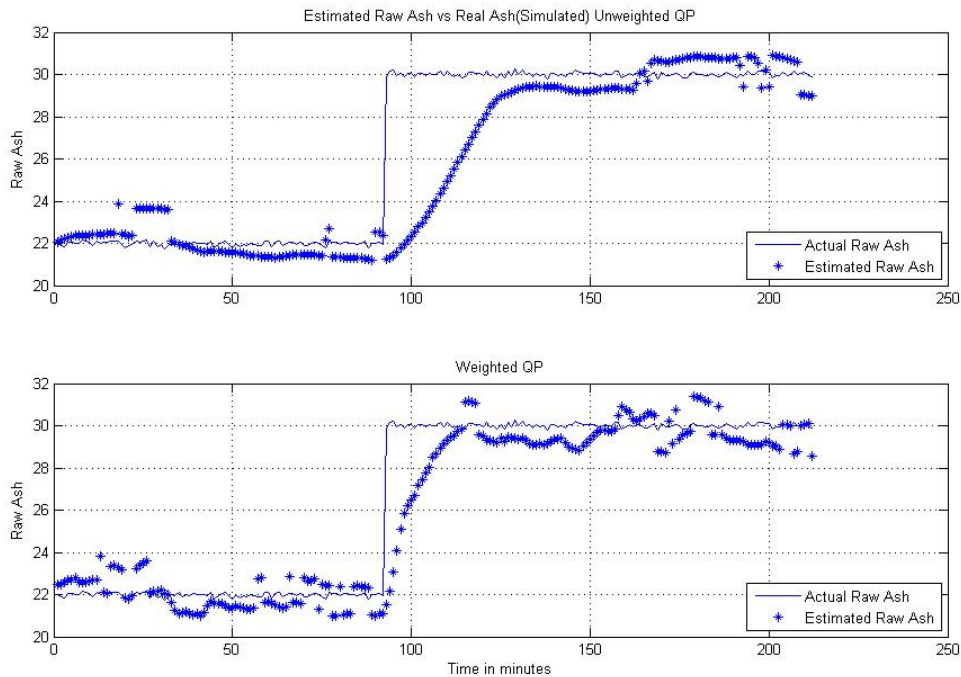


Figure 2 Weighted vs Unweighted QP's

This dynamic QP method for blending barges was implemented at a coal mine for the purpose of blending raw and clean coal to a high ash specification on 1000-1500 ton barges where the purpose was to maximize the amount of raw on each barge. The method worked extremely well for three years until the end of the mines life.

- [1] N.S. Nies. *Control Systems Engineering*, Benjamin/Cummings Publishing Co, 1992.
- [2] J.V. De Vegte, *Feedback Control Systems*, Prentice-Hall, (2nd edition), 1990.
- [3] J.M. Maciejowski. *Predictive Control With Constraints*. Prentice Hall, 2002.
- [4] E.F Camacho, C. Bordons. *Model Predictive Control*, Springer, (2nd edition), 2007.
- [5] B.A.Bancroft, G.R.Hobbs. "Distribution of the Kriging Error and Stationarity of the Variogram in a Coal Property," *Mathematical Geology*, vol. 18, no. 7, 1986.
- [6] A.S.Goldberger. "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *J. Am. Statis. Assoc.*, vol.57, 369-375, 1962.

Bruce A. Bancroft, CONSOL Energy Inc.